

Learning Rate Decay

Intuitive: we want to decrease learning rate as we approach the minimum otherwise our model would zigzag around the minimum.

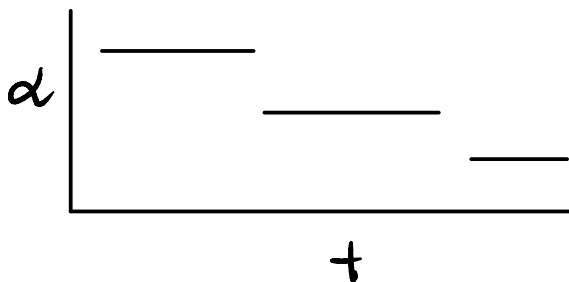
1 epoch = 1 pass through data.

$$\alpha = \frac{1}{1 + \text{decay-rate} \times \text{epoch-num}} \alpha_0.$$

Other methods:

$$\alpha = 0.95^{\text{epoch-num}} \cdot \alpha_0 \quad \text{--- exponential decay.}$$

$$\alpha = \frac{k}{\sqrt{\text{epoch-num}}} \cdot \alpha_0 \quad \text{or} \quad \frac{k}{\sqrt{t}} \cdot \alpha_0$$



discrete staircase.

Manual Decay.